(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 November 2002 (21.11.2002)

PCT

(10) International Publication Number
WO 02/092780 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/15767

(22) International Filing Date: 17 May 2002 (17.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/300,381	17 May 2001 (17.05.2001)	US
60/300,907	25 June 2001 (25.06.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/300,907 (CIP)
Filed on	25 June 2001 (25.06.2001)
US	60/300,381 (CIP)
Filed on	17 May 2001 (17.05.2001)

(71) Applicant *(for all designated States except US)*: **DIVERSA CORPORATION** [US/US]; 4955 Directors Place, San Diego, CA 92121 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **SHORT, Jay, M.** [US/US]; P.O. Box 7214, Rancho Santa Fe, CA 92067-7214 (US).

(74) Agent: **EINHORN, Gregory, P.**; Fish & Richardson P.C., 4350 La Jolla Village Drive, Suite 500, San Diego, CA 92122 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/092780 A2

(54) Title: NOVEL ANTIGEN BINDING MOLECULES FOR THERAPEUTIC, DIAGNOSTIC, PROPHYLACTIC, ENZYMATIC, INDUSTRIAL, AND AGRICULTURAL APPLICATIONS, AND METHODS FOR GENERATING AND SCREENING THEREOF

(57) Abstract: The invention is directed to methods for generating sets, or libraries, of nucleic acids encoding antigen-binding sites, such as antibodies, antibody domains or other fragments, including single and double stranded antibodies, major histocompatibility complex (MHC) molecules, T cell receptors (TCRs), and the like. This invention provides methods for generating variant antigen binding sites, e.g., antibodies and specific domains or fragments of antibodies (e.g., Fab or Fc domains), by altering template nucleic acids including by saturation mutagenesis, synthetic ligation reassembly, or a combination thereof. In one aspect, the invention provides methods for generating all human or humanized antibodies and evolving them to achieve optimized properties related to stability, duration, expression, production, enzymatic activity, affinity, avidity, localization, and other immunological properties. Polypeptides generated by these methods can be analyzed using a novel capillary array platform, which provides unprecedented ultra-high throughput screening.

NOVEL ANTIGEN BINDING MOLECULES FOR THERAPEUTIC, DIAGNOSTIC, PROPHYLACTIC, ENZYMATIC, INDUSTRIAL, AND AGRICULTURAL APPLICATIONS, AND METHODS FOR GENERATING AND SCREENING THEREOF

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims the benefit of priority under 35 U.S.C. §119(e) of U.S. Provisional Application Nos. 60/300,381, filed May 17, 2001, and 60/300,907, filed June 25, 2001; and is a continuation-in-part (CIP) of U.S. Patent Application Serial No. (USSN) 09/535,754, filed March 27, 2000 (entitled Exonuclease-Mediated Gene Assembly in Directed Evolution); which is a CIP of USSN 09/522,289, filed March 9, 2000 (entitled End Selection in Directed Evolution); which is a CIP of USSN 09/498,557, filed February 4, 2000 (entitled Non-Stochastic Generation of Genetic Vaccines and Enzymes), which is hereby incorporated by reference; which is a CIP of USSN 09/495,052, filed on January 31, 2000 (entitled Non-Stochastic Generation of Genetic Vaccines); which is a CIP of USSN 09/276,860, filed on March 26, 1999 (entitled Exonuclease-Mediated Gene Assembly in Directed Evolution); which is a CIP of USSN 09/267,118, filed on March 9, 1999 (entitled End Selection in Directed Evolution); which is a continuation-in part of USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution, now USPN 6,171,820); which is a continuation of USSN 09/185,373 filed on November 3, 1998 (entitled Directed Evolution of Thermophilic Enzymes); which is a continuation of USSN 08/760,489 filed on Dec. 5, 1996 (entitled Directed Evolution of Thermophilic Enzymes, now USPN 5,830,696); which claims the benefit of U.S. provisional application number 60/008,311 filed on Dec. 07, 1995.

USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution) is also a CIP of USSN 08/962,504 filed on October 31, 1997 (entitled Method of DNA Shuffling); which is a CIP of USSN 08/677,112 filed on July 9, 1996 (entitled Method of DNA Reassembly by Interrupting Synthesis, now USPN 5,965,408).

USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution) is also a CIP of USSN 08/651,568 filed on May 22, 1996 (entitled Production of Enzymes Having Desired Activities by Mutagenesis, now USPN 5,939,250); which claims the

benefit of U.S. provisional application serial No. 60/008,316, filed December 7, 1995 (entitled Combinatorial Enzyme Development).

The present application is also a CIP of PCT application No. PCT/US00/16838, filed June 14, 2000 (entitled Synthetic Ligation Reassembly in Directed Evolution, now PCT publication No. WO 00/77262); which claims the benefit of USSN 09/594,459, filed June 14, 2000 (entitled Synthetic Ligation Reassembly in Directed Evolution); which is a CIP of USSN 09/332,835, filed June 14, 1999 (entitled Synthetic Ligation Reassembly in Directed Evolution).

The present application is also a CIP of PCT application No. PCT/US00/08245, filed March 27, 2000 (entitled Exonuclease-Mediated Nucleic Acid Reassembly in Directed Evolution, now PCT publication No. WO 00/58517); which claims the benefit of USSN 09/276,860, filed on March 26, 1999 (entitled Exonuclease-Mediated Gene Assembly in Directed Evolution); which is a CIP of USSN 09/267,118, filed on March 9, 1999 (entitled End Selection in Directed Evolution); which is a continuation-in part of USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution, now USPN 6,171,820); which is a continuation of USSN 09/185,373 filed on November 3, 1998 (entitled Directed Evolution of Thermophilic Enzymes); which is a continuation of USSN 08/760,489 filed on Dec. 5, 1996 (entitled Directed Evolution of Thermophilic Enzymes, now USPN 5,830,696); which claims the benefit of U.S. provisional application number 60/008,311 filed on Dec. 07, 1995.

The present application is also a CIP of PCT application No. PCT/US00/06497, filed March 9, 1999 (entitled End Selection in Directed Evolution, now PCT publication No. WO 00/53744); which claims the benefit of USSN 09/332,835, filed June 14, 1999 (entitled Synthetic Ligation Reassembly in Directed Evolution). PCT application No. PCT/US00/06497, filed March 9, 1999 (entitled End Selection in Directed Evolution, now PCT publication No. WO 00/53744) also claims the benefit of USSN 09/267,118, filed on March 9, 1999 (entitled End Selection in Directed Evolution); which is a continuation-in part of USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution, now USPN 6,171,820); which is a continuation of USSN 09/185,373 filed on November 3, 1998 (entitled Directed Evolution of Thermophilic Enzymes); which is a continuation of USSN 08/760,489 filed on Dec. 5, 1996 (entitled Directed Evolution of Thermophilic Enzymes, now USPN 5,830,696); which claims the benefit of U.S. provisional application number 60/008,311 filed on Dec. 07, 1995.

PCT application No. PCT/US00/06497, filed March 9, 1999 (entitled End Selection in Directed Evolution, now PCT publication No. WO 00/53744) also claims the benefit of USSN

09/276,860, filed on March 26, 1999 (entitled Exonuclease-Mediated Gene Assembly in Directed Evolution); which is a CIP of USSN 09/267,118, filed on March 9, 1999 (entitled End Selection in Directed Evolution); which is a continuation-in part of USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution, now USPN 6,171,820); which is a continuation of USSN 09/185,373 filed on November 3, 1998 (entitled Directed Evolution of Thermophilic Enzymes); which is a continuation of USSN 08/760,489 filed on Dec. 5, 1996 (entitled Directed Evolution of Thermophilic Enzymes, now USPN 5,830,696); which claims the benefit of U.S. provisional application number 60/008,311 filed on Dec. 07, 1995.

The present application is also a CIP of USSN 09/594,459, filed June 14, 2000 (entitled Synthetic Ligation Reassembly in Directed Evolution); which is a CIP of USSN 09/332,835, filed June 14, 1999 (entitled Synthetic Ligation Reassembly in Directed Evolution).

The present application is also a CIP of PCT application No. PCT/US00/03086, filed February 4, 2000 (entitled Non-Stochastic Generation of Genetic Vaccines and Enzymes); which claims the benefit of USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution, now USPN 6,171,820); which is a continuation of USSN 09/185,373 filed on Nov. 3, 1998 (entitled Directed Evolution of Thermophilic Enzymes); which is a continuation of USSN 08/760,489 filed on Dec. 5, 1996 (entitled Directed Evolution of Thermophilic Enzymes, now USPN 5,830,696); which claims the benefit of U.S. provisional application number 60/008,311 filed on Dec. 07, 1995.

The present application is also a CIP of USSN 09/756,459, filed January 8, 2001 (entitled Saturation Mutagenesis in Directed Evolution); which is a continuation of USSN 09/246,178, filed Feb. 4, 1999 (entitled Saturation Mutagenesis in Directed Evolution, now USPN 6,171,820); which is a continuation of USSN 09/185,373 filed on Nov. 3, 1998 (entitled Directed Evolution of Thermophilic Enzymes); which is a continuation of USSN 08/760,489 filed on Dec. 5, 1996 (entitled Directed Evolution of Thermophilic Enzymes, now USPN 5,830,696); which claims the benefit of U.S. provisional application number 60/008,311 filed on Dec. 07, 1995.

The present application is also a CIP of USSN [UNASSIGNED], filed January 9, 2001 (entitled Optimized Directed Evolution System and Method).

The present application is also a CIP of USSN 09/376,727, filed August 17, 1999 (entitled Method of DNA Shuffling with Polynucleotides Produced by Blocking or Interrupting a Synthesis or Amplification Process); which is a continuation of USSN 08/677,112, filed July 9, 1996 (entitled Method of DNA Reassembly by Interrupting Synthesis, now USPN 5,965,408).

The present application is also a CIP of PCT application No. PCT/US98/22596, filed October 23, 1998 (entitled Method of DNA Shuffling); which claims the benefit of USSN 09/962,504, filed October 31, 1997 (entitled Method of DNA Shuffling); which is a CIP of USSN 08/677,112, filed July 9, 1996 (entitled Method of DNA Reassembly by Interrupting Synthesis, now USPN 5,965,408).

The present application is also a CIP of USSN 09/214,645, filed September 27, 1999 (entitled Method of DNA Shuffling with Polynucleotides Produced by Blocking or Interrupting a Synthesis or Amplification Process); which is a national phase application of PCT application No. PCT/US97/12239, filed July 9, 1997 (entitled Method of DNA Shuffling with Polynucleotides Produced by Blocking or Interrupting a Synthesis or Amplification Process, now PCT publication No. WO 98/01581); which claims the benefit of USSN 08/677,112, filed July 9, 1996 (entitled Method of DNA Reassembly by Interrupting Synthesis, now USPN 5,965,408).

The present application is also a CIP of USSN 09/790,321, filed February 21, 2001 (entitled Capillary Array-Based Enzyme Screening); which is a divisional of USSN 09/687,219, filed October 12, 2000 (entitled Capillary Array-Based Sample Screening); which is a CIP of USSN 09/636,778, filed August 11, 2000 (entitled High Throughput Screening of Novel Enzymes); which is a continuation of USSN 09/098,206, filed June 16, 1998 (entitled High Throughput Screening of Novel Enzymes, now USPN 6,174,673); which is a CIP of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

The present application is also a CIP of USSN 09/761,559, filed January 16, 2001 (entitled High Throughput Screening of Novel Enzymes); which is a divisional of USSN 09/098,206, filed June 16, 1998 (entitled High Throughput Screening of Novel Enzymes, now USPN 6,174,673); which is a CIP of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

The present application is also a CIP of USSN 09/848,185 filed May 3, 2001 (entitled High Throughput Screening for Novel Enzymes); which is a divisional of USSN 09/636,778, filed August 11, 2000 (entitled High Throughput Screening of Novel Enzymes); which is. a continuation of USSN 09/098,206, filed June 16, 1998 (entitled High Throughput Screening of Novel Enzymes, now USPN 6,174,673); which is a CIP of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

The present application is also a CIP of USSN 09/738,871, filed December 14, 2000 (entitled High Throughput Screening for a Bioactivity or Biomolecule); which is a CIP of USSN

09/685,432, filed October 10, 2000 (entitled High Throughput Screening for Sequences of Interest); which is a CIP of USSN 09/444,112, filed November 22, 1999 (entitled Capillary Array-Based Enzyme Screening); which is a CIP of USSN 09/098,206, filed June 16, 1998 (entitled High Throughput Screening of Novel Enzymes, now USPN 6,174,673); which is a CIP of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

The present application is also a CIP of PCT application No. PCT/US00/32208, filed November 22, 2000 (entitled Capillary Array-Based Sample Screening); which claims the benefit of USSN 09/687,219, filed October 12, 2000 (entitled Capillary Based-Based Sample Screening); which is a CIP of USSN 09/636,778, filed August 11, 2000 (entitled High Throughput Screening of Novel Enzymes); which is a continuation of USSN 09/098,206, filed June 16, 1998 (entitled High Throughput Screening of Novel Enzymes, now USPN 6,174,673); which is a CIP of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

The present application is also a CIP of PCT application No. PCT/US98/12674, filed June 16, 1998 (entitled High Throughput Screening for Novel Enzymes, now PCT publication No. WO 98/58085); which claims the benefit of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

PCT/US00/32208, filed November 22, 2000 (entitled Capillary Array-Based Sample Screening), also claims the benefit of USSN 09/444,112, filed November 22, 1999 (entitled Capillary Array-Based Enzyme Screening); which is a CIP of USSN 09/098,206, filed June 16, 1998 (entitled High Throughput Screening of Novel Enzymes, now USPN 6,174,673); which is a CIP of USSN 09/876,276, filed June 16, 1997 (entitled High Throughput Screening of Novel Enzymes).

These aforementioned applications and patents are explicitly incorporated herein by reference in their entirety and for all purposes.

TECHNICAL FIELD

The present invention is generally directed to the fields of medicine, protein engineering, immunology and molecular biology. In one aspect, the invention is directed to methods for generating sets, or libraries, of nucleic acids encoding antigen binding molecules, including, e.g., antibodies and related molecules, such as antigen binding sites and domains and other antigen binding fragments, including single and double stranded antibodies, T cell receptors (TCRs) and Class I and Class II major histocompatibility (MHC) molecules. This invention also provides methods for generating new or variant antigen binding polypeptides, e.g., antigen binding sites, antibodies and specific domains or fragments of antibodies (e.g., Fab or Fc domains), TCRs and MHC molecules by altering template nucleic acids by, e.g., saturation mutagenesis, an optimized directed evolution system, synthetic ligation reassembly, or a combination thereof.

Polypeptides generated by these methods can be analyzed using any liquid or solid state screening method, e.g., phage display, ribosome display, using capillary array platforms, and the like. The polypeptides generated by the methods of the invention can be used *in vitro*, e.g., to isolate or identify antigens or *in vivo*, e.g., to treat or diagnose various diseases and conditions, to modulate, stimulate or attenuate an immune response.

This invention pertains to the field of genetic vaccines. Specifically, the invention provides multi-component genetic vaccines that contain components that are optimized for a particular vaccination goal. In one aspect, this invention provides methods for improving the efficacy of genetic vaccines by providing materials that facilitate targeting of a genetic vaccine to a particular tissue or cell type of interest. The invention also provides antigen binding molecules, e.g., T cell receptors and Class I and Class II major histocompatibility (MHC) molecules, having an engineered affinity to an antigen, thus allowing manipulation of the immune response to the vaccine.

This invention pertains to the field biologic therapeutics by providing polypeptides comprising antigen binding sites, such as antibodies, with modified (e.g., increased or decreased) affinity for antigen. For example, the methods of the invention provide antibodies of altered or enhanced affinities for an antigen for use, e.g., in immunotherapeutics or diagnostics. The antibodies generated by the methods of the invention can be administered therapeutically to slow the growth of or kill cells, such as cancer cells, or, to stimulate cell division, e.g., for enhancing an immune response or for

tissue regeneration, or, to alter any biological mechanism or response. For example, administration of antibodies that bind to immune effector or regulatory cells, or to lymphokines or cytokines, can alter, e.g., upregulate, stimulate or attenuate, an humoral or a cellular immune response.

This invention pertains to the field of modulation of immune responses such as those induced by genetic vaccines and also pertains to the field of methods for developing immunogens that can induce efficient immune responses against a broad range of antigens.

This invention pertains to the field of modulation of immune responses by modifying molecules that are involved in the stimulation and regulation of the immune response, including, e.g., T cell receptors and Class I and Class II major histocompatibility (MHC) molecules. Thus, molecules generated by the methods of the invention can have increased or decreased affinity of binding sites to antigen. For example, by decreasing the affinity of a T cell receptor for an antigen (which a TCR binds in conjunction with an MHC molecule, i.e., the MHC "presents" the antigen to the TCR), the methods of the invention can generate a non-autoreactive variant of an autoreactive TCR. In another example, by increasing the affinity of an MHC molecule for an antigen, e.g., a pathogenic antigen, the methods of the invention can generate an enhanced immune response to that pathogen. Similarly, if the antigen is a self antigen, by decreasing the affinity of the MHC molecule for the antigen, the methods of the invention can generate an abated or attenuated immune response to that self antigen.

Thus, the present invention also relates generally to novel proteins, and fragments thereof, as well as nucleic acids which encode these proteins, and methods of making and using these proteins in diagnostic, prophylactic and therapeutic applications. In a particular exemplification, the present invention relates to proteins from the *Plasmodium falciparum* erythrocyte membrane protein 1 ("PfEMP1") gene family and fragments thereof which are derived from malaria-parasitized erythrocytes. In particular, these proteins are derived from the erythrocyte membrane protein of *Plasmodium falciparum* parasitized erythrocytes, also termed "PfEMP1". The present invention also provides nucleic acids encoding these proteins, which proteins and nucleic acids are associated with the pathology of malaria infections, and which may be used as vaccines or other prophylactic treatments for the prevention of malaria infections, and/or in diagnosing and treating the symptoms of patients who suffer from malaria and associated diseases.

This invention also relates to the field of protein engineering. Specifically, this invention relates to a directed evolution method for preparing a polynucleotide encoding a polypeptide. More specifically, this invention relates to a method of using mutagenesis to generate a novel polynucleotide encoding a novel polypeptide, which novel polypeptide is itself an altered ("improved") biological molecule and/or contributes to the generation of another improved biological molecule. More specifically still, this invention relates to a method of performing both non-stochastic polynucleotide chimerization and non-stochastic site-directed point mutagenesis.

Thus, in one aspect, this invention relates to a method of generating a progeny library, or set, of chimeric polynucleotide(s) by means that are synthetic and non-stochastic. The design of the progeny polynucleotide(s) is derived by analysis of a parental set of polynucleotides and/or of the polypeptides correspondingly encoded by the parental polynucleotides. In another aspect, this invention relates to a method of performing site-directed mutagenesis using means that are exhaustive, systematic, and non-stochastic.

Furthermore this invention relates to a step of selecting from among a generated set of progeny molecules a subset comprised of particularly desirable species, including by a process termed end-selection, which subset may then be screened further. This invention also relates to the step of screening a set of polynucleotides for the production of a polypeptide and/or of another expressed biological molecule having a useful property, such as an antibody with increased affinity for an antigen.

Novel biological molecules whose manufacture is taught by this invention include genes, gene pathways, and any molecules whose expression is affected thereby, including directly encoded polypetides and /or any molecules affected by such polypeptides. Said novel biological molecules include those that contain a carbohydrate, a lipid, a nucleic acid, and /or a protein component, and specific but non-limiting examples of these include antibiotics, antibodies, TCRs, MHC molecules, enzymes, and steroidal and non-steroidal hormones.

In one aspect, the present invention relates to enzymes, particularly to thermostable enzymes, and to their generation by directed evolution. More particularly, the present invention relates to thermostable enzymes which are stable at high temperatures and which have improved activity at lower temperatures.

BACKGROUND

Antigen binding polypeptides, such as antibodies, are increasingly used in a variety of therapeutic applications. For example, in immunotherapy, antibodies are used to directly kill target cells, such as cancer cells. Antigen binding polypeptides are also used as carriers to deliver cytotoxic or imaging reagents. Monoclonal antibodies (mAbs) approved for cancer therapy are now in Phase II and III trials. Certain anti-idiotypic antibodies that bind to the antigen-combining sites of antibodies can effectively mimic the three-dimensional structures and functions of the external antigens and can be used as surrogate antigens for active specific immunotherapy. Bi-specific antibodies combine immune cell activation with tumor cell recognition; thus, tumor cells or cells expressing tumor specific antigens (e.g., tumor vasculature) are killed by pre-defined effector cells. Antibodies can be administered to increase or decrease the levels of cytokines or hormones by direct binding or by stimulating or inhibiting secretory cells. Accordingly, increasing the affinity or avidity of an antibody to a desired antigen, such as a cancer-specific antigen, would result in greater specificity of the antibody to its target, resulting in a variety of therapeutic benefits, such as needing to administer less antibody-containing pharmaceutical.

Providing protective immunity even in situations when the pathogens are poorly characterized or cannot be isolated or cultured in laboratory environment.

Genetic immunization represents a novel mechanism of inducing protective humoral and cellular immunity. Vectors for genetic vaccinations generally consist of DNA that includes a promoter/enhancer sequence, the gene of interest and a polyadenylation/ transcriptional terminator sequence. After intramuscular or intradermal injection, the gene of interest is expressed, followed by recognition of the resulting protein by the cells of the immune system. Genetic immunizations provide means to induce protective immunity even in situations when the pathogens are poorly characterized or cannot be isolated or cultured in laboratory environment.

Small improvement in the efficiency of genetic vaccine vectors can result in dramatic increase if the level of immune response

The efficacy of genetic vaccination is often limited by inefficient uptake of genetic vaccine vectors into cells. Generally, less than 1% of the muscle or skin cells at the sites of injections express the gene of interest. Even a small improvement in the efficiency of genetic vaccine vectors to enter the cells can result in a dramatic increase in the level of immune

response induced by genetic vaccination. A vector typically has to cross many barriers which can result in only a very minor fraction of the DNA ever being expressed.

Various limitations to immunogenicity

Limitations to immunogenicity include: loss of vector due to nucleases present in blood and tissues; inefficient entry of DNA into a cell; inefficient entry of DNA into the nucleus of the cell and preference of DNA for other compartments; lack of DNA stability in the nucleus (factor limiting nuclear stability may differ from those affecting other cellular and extracellular compartments), and, for vectors that integrate into the chromosome, the efficiency of integration and the site of integration. Moreover, for many applications of genetic vaccines, it is preferable for the genetic vaccine to enter a particular target tissue or cell.

Thus, a need exists for genetic vaccines that can be targeted to specific cell and tissue types of interest, and which exhibit an increased ability to enter the target cells.

Pathways for immune responses induced by genetic vaccines

Elicitation of a desired *in vivo* response by a genetic vaccine generally requires multiple cellular processes in a complex sequence. Several potential pathways exist along which a genetic vaccine can exert its effect on the mammalian immune system. In one pathway, the genetic vaccine vector enters cells that are the predominant cell type in the tissue that receives vaccine (e.g., muscle or epithelial cells). These cells express and release the antigen encoded by the vector. The vaccine vector can be engineered to have the antigen released as an intact protein from living transfected cells (i.e., via a secretion process) or directed to a membrane-bound form on the surface of these cells. Antigen can also be released from an intracellular compartment of such cells if those cells die.

The antigen derived from vaccine vector internalization and antigen expression within the predominant cell type in the tissue ends up within APC, which then process the antigen internally to prime MHC Class I and or Class II, essential steps in activation of $CD4^+$ T-helper cells and development of potent specific immune responses.

Extracellular antigen derived from any of these situations interacts with antigen presenting cells (APC) either by binding to the cell surface (specifically via IgM or via other non-immunoglobulin receptors) and subsequent endocytosis of outer membrane, or by fluid phase micropinocytosis wherein the APC internalizes extracellular fluid and its contents into an endocytic compartment. Interaction with APC may occur before or after partial proteolytic

cleavage in the extracellular environment. In any case, the antigen derived from vaccine vector internalization and antigen expression within the predominant cell type in the tissue ends up within APC. The APC then process the antigen internally to prime MHC Class I and or Class II, essential steps in activation of $CD4^+$ T-helper cells (T_H1 and/or T_H2) and development of potent specific immune responses.

The genetic vaccine plasmid enters APC and antigen is proteolytically cleaved in the cell cytoplasm.

In a parallel pathway, the genetic vaccine plasmid enters APC (or the predominant cell type in the tissue) and, instead of antigen derived from plasmid expression being directed to extracellular export, antigen is proteolytically cleaved in the cell cytoplasm (in a proteasome dependent or independent process). Often, intracellular processing in such cells occurs via proteasomal degradation into peptides that are recognized by the TAP-1 and TAP-2 proteins and transported into the lumen of the rough endoplasmic reticulum (RER).

The peptide fragments are transported into the RER complex, expressed on the cell surface; in the presence of appropriate additional signals, can differentiate into functional CTLs.

The peptide fragments transported into the RER complex with MHC Class I. Such antigen fragments are then expressed on the cell surface in association with Class I. $CD8^+$ cytotoxic T lymphocytes (CTL) bearing specific T cell receptor then recognize the complex and can, in the presence of appropriate additional signals, differentiate into functional CTLs.

By virtue of poorly characterized pathways for trafficking of cytoplasmically generated peptides into endosomal compartments, a genetic vaccine vector can lead to $CD4^+$ T cell stimulation.

In addition, poorly characterized pathways, which are generally not dominant, exist in APC for trafficking of cytoplasmically generated peptides into endosomal compartments where they can end up complexed with MHC Class II, and thereby act to present antigen peptides to $CD4^+$ T_H1 and T_H2 cells. Because activation, proliferation, differentiation and immunoglobulin isotype switching by B lymphocytes requires help of $CD4^+$ T cells, antigen presentation in the context of MHC Class II molecules is crucial for induction of antigen-specific antibodies. By virtue of this pathway, a genetic vaccine vector can lead to $CD4^+$ T cell stimulation in addition to the dominant $CD8^+$ CTL activation process described above. This alternative pathway is, however, of little consequence in muscle cells where levels of MHC Class II expression are very low or zero.

In this case cytokines are derived not only from processes intrinsic to the interaction of DNA with cells, or specific cell responses to the antigen, but via synthesis directed by the vaccine plasmid.

Genetic vaccination can also elicit cytokine release from cells that bind to or take up DNA. So-called immunostimulatory or adjuvant properties of DNA are derived from its interaction with cells that internalize DNA. Cytokines can be released from cells that bind and/or internalize DNA in the absence of gene transcription. Separately, interaction of antigen with APC followed by presentation and specific recognition also stimulates release of cytokines that have positive feedback effects on these cells and other immune cells. Chief among these effects are the direction of $CD4^+$ T_H cells to differentiate/ proliferate preferentially to T_H1 or T_H2 phenotypes. Furthermore, cytokines released at the site of DNA vaccination, regardless of the mechanism of their release, contribute to recruitment of other immune cells from the immediate local area and more distant sites such as draining lymph nodes. In recognition of the importance of cytokines in elicitation of a potent immune response, some investigators have included the genes for one or more cytokines in the DNA vaccine plasmid along with the target antigen for immunization. In this case cytokines are derived not only from processes intrinsic to the interaction of DNA with cells, or specific cell responses to the antigen, but via synthesis directed by the vaccine plasmid.

Movement of immune cells from the blood stream and different sites to the site of immunization and also from the site of immunization to other sites

Immune cells are recruited to the site of immunization from distant sites or the bloodstream. Specific and non-specific immune responses are then greatly amplified. Immune cells, including APC, bearing antigen fragments complexed to MHC molecules or even expressing antigen from uptake of plasmid, also move from the immunization site to other sites (blood, hence to all tissues; lymph nodes; spleen) where additional immune recruitment and qualitative and quantitative development of the immune response ensue.

Current genetic vaccine vectors employ simple methods for expression of the desired antigen with few if any design elements that control the precise intracellular fate of the antigen or the immunological consequences of antigen expression

While these pathways often compete, previously available genetic vaccines have incorporated all components for influencing each of the pathways into a single polynucleotide molecule. Because separate cell types are involved in the complex

interactions required for a potent immune response to a genetic vaccine vector, mutually incompatible consequences can arise from administration of a genetic vaccine that is incorporated in a single vector molecule. Current genetic vaccine vectors employ simple methods for expression of the desired antigen with few if any design elements that control the precise intracellular fate of the antigen or the immunological consequences of antigen expression. Thus, although genetic vaccines show great promise for vaccine research and development, the need for major improvements and several severe limitations of these technologies are apparent.

Existing genetic vaccine vectors have not been optimized for human tissue, providing low and short-lasting expression of the antigen of interest, with insufficient stability, inducibility, or levels of expression *in vivo*, among other things

Largely due to the lack of suitable laboratory models, none of the existing genetic vaccine vectors have been optimized for human tissues. The existing genetic vaccine vectors typically provide low and short-lasting expression of the antigen of interest, and even large quantities of DNA do not always result in sufficiently high expression levels to induce protective immune responses. Because the mechanisms of the vector entry into the cells and transfer into the nucleus are poorly understood, virtually no attempts have been made to improve these key properties. Similarly, little is known about the mechanisms that regulate the maintenance of vector functions, including gene expression. Furthermore, although there is increasing amount of data indicating that specific sequences alter the immunostimulatory properties of the DNA, rational engineering is a very laborious and time- consuming approach when using this information to generate vector backbones with improved immunomodulatory properties.

Moreover, presently available genetic vaccine vectors do not provide sufficient stability, inducibility or levels of expression *in vivo* to satisfy the desire for vaccines which can deliver booster immunization without additional vaccine administration. Booster immunizations are typically required 3-4 weeks after the primary injection with existing genetic vaccines. Therefore a need exists for improved genetic vaccine vectors and formulations, and methods for development of such vectors.

The interactions between pathogens and hosts are results of millions of years of evolution, during which the mammalian immune system has evolved sophisticated means to counterattack pathogen invasions. However, bacterial and viral pathogens have

simultaneously gained a number of mechanisms to improve their virulence and survival in hosts, providing a major challenge for vaccine research and development despite the powers of modem techniques of molecular and cellular biology. Similar to the evolution of pathogen antigens, several cancer antigens are likely to have gained means to downregulate their immunogenicity as a mechanism to escape the host immune system.

Efficient vaccine development is also hampered by the antigenic heterogeneity of different strains of pathogens, driven in part by evolutionary forces as means for the pathogens to escape immune defenses. Pathogens also reduce their immunogenicity by selecting antigens that are difficult to express, process and/or transport in host cells, thereby reducing the availability of immunogenic peptides to the molecules initiating and modulating immune responses. The mechanisms associated with these challenges are complex, multivariate and rather poorly characterized. Accordingly, a need exists for vaccines that can induce a protective immune response against bacterial and viral pathogens.

Antigen processing and presentation is only one factor which determines the effectiveness of vaccination, whether performed with genetic vaccines or more classical methods. Other molecules involved in determining vaccine effectiveness include cytokines (interleukins, interferons, chemokines, hematopoietic growth factors, tumor necrosis factors and transforming growth factors), which are small molecular weight proteins that regulate maturation, activation, proliferation and differentiation of the cells of the immune system.

Characteristic features of cytokines are pleiotropy and redundancy; that is, one cytokine often has several functions and a given function is often mediated by more than one cytokine. In addition, several cytokines have additive or synergistic effects with other cytokines, and a number of cytokines also share receptor components.

Due to the complexity of the cytokine networks, studies on the physiological significance of a given cytokine have been difficult, although recent studies using cytokine gene-deficient mice have significantly improved our understanding on the functions of cytokines *in vivo*. In addition to soluble proteins, several membrane- bound costimulatory molecules play a fundamental role in the regulation of immune responses. These molecules include CD40, CD40 ligand, CD27, CD80, CD86 and CD150 (SLAM), and they are typically expressed on lymphoid cells after activation via antigen recognition or through cell-cell interactions.

T helper (T_H) cells, key regulators of the immune system, are capable of producing a large number of different cytokines, and based on their cytokine synthesis pattern T_H cells are divided into two subsets (Paul and Seder (1994) *Cell* 76: 241-251). T_H1 cells produce high levels of IL-2 and IFN-γ and no or minimal levels of IL-4, IL-5 and IL-13. In contrast, T_H2 cells produce high levels of IL-4, IL-5 and IL-13, and IL-2 and IFN-γ production is minimal or absent. T_H1 cells activate macrophages, dendritic cells and augment the cytolytic activity of $CD8^+$ cytotoxic T lymphocytes and NK cells (*Id.*), whereas T_H2 cells provide efficient help for B cells and they also mediate allergic responses due to the capacity of T_H2 cells to induce IgE isotype switching and differentiation of B cells into IgE secreting cell (De Vries and Punnonen (1996) In *Cytokine regulation of humoral immunity: basic and clinical aspects*. Eds. Snapper, C.M., John Wiley & Sons, Ltd., West Sussex, UK, p. 195- 215). The exact mechanisms that regulate the differentiation of T helper cells are not fully understood, but cytokines are believed to play a major role. IL-4 has been shown to direct T_H2 differentiation, whereas IL-12 induces development of T_H1 cells (Paul and Seder, *supra.*). In addition, it has been suggested that membrane bound costimulatory molecules, such as CD80, CD86 and CD150, can direct T_H1 and/or T_H2 development, and the same molecules that regulate T_H cell differentiation also affect activation, proliferation and differentiation of B cells into Ig-secreting plasma cells (Cocks *et al.* (1995) *Nature* 376: 260-263; Lenschow *et al.* (1996) *Immunity* 5: 285-293; Punnonen *et al.* (1993) *Proc. Nat'l. Acad. Sci. USA* 90: 3730-3734; Punnonen *et al.* (1997) *J Exp. Med.* 185: 993-1004).

Studies in both man and mice have demonstrated that the cytokine synthesis profile of T helper (T_H) cells plays a crucial role in determining the outcome of several viral, bacterial and parasitic infections. High frequency of T_H1 cells generally protects from lethal infections, whereas dominant T_H2 phenotype often results in disseminated, chronic infections. For example, T_HI phenotype is observed in tuberculoid (resistant) form of leprosy and T_H2 phenotype in lepromatous, multibacillary (susceptible) lesions (Yamamura *et al.* (199 1) *Science* 254: 277-279). Similarly, late-stage HIV patients have T_H2-like cytokine synthesis profiles, and T_H1 phenotype has been proposed to protect from AIDS (Maggi *et al.* (1994) *J Exp. Med.* 180: 489-495). Furthermore, the survival from meningococcal septicemia is genetically determined based on the capacity of peripheral blood leukocytes to produce TNF-α and IL- 10. Individuals from families with high production of IL- 10 have increased risk of fatal meningococcal disease, whereas members of families with high TNF- α

production were more likely to survive the infection (Westendorp *et al.* (1997) *Lancet* 349: 170-173).

Cytokine treatments can dramatically influence T_H1/T_H2 cell differentiation and macrophage activation, and thereby the outcome of infectious diseases. For example, BALB/c mice infected with *Leishmania major* generally develop a disseminated fatal disease with a T_H2 phenotype, but when treated with anti-IL-4 mAbs or IL- 12, the frequency of T_H1 cells in the mice increases and they are able to counteract the pathogen invasion (Chatelain *et al.* (1992) *J Immunol.* 148: 1182-1187). Similarly, IFN-γ protects mice from lethal Herpes Simplex Virus (HSV) infection, and MCP-1 prevents lethal infections by *Pseudomonas aeruginosa* or *Salmonella typhimurium.* In addition, cytokine treatments, such as recombinant IL-2, have shown beneficial effects in human common variable immunodeficiency (Cunningham-Rundles *et al.* (1994) *N. Engl. J Med.* 331: 918-921).

The administration of cytokines and other molecules to modulate immune responses in a manner most appropriate for treating a particular disease can provide a significant tool for the treatment of disease. However, presently available immunomodulator treatments can have several disadvantages, such as insufficient specific activity, induction of immune responses against, the immunomodulator that is administered, and other potential problems. Thus, a need exists for immunomodulators that exhibit improved properties relative to those currently available.

Erythrocytes infected with the malaria parasite *P. falciparum* disappear from the peripheral circulation as they mature from the ring stage to trophozoites (Bignami and Bastianeli, *Reforma Medica* (1889) 6:1334-1335). This phenomenon, known as sequestration, results from parasitized erythrocyte ("PE") adherence to microvascular endothelial cells in diverse organs (Miller, *Am. J. Trop. Med. Hyg.* (1969) 18:860-865). Sequestration is associated temporally with expression of knob protrusions (Leech *et al., J. Cell. Biol.* (1984) 98:1256-1264), expression of a very large antigenically variant surface protein, called PfEMP1 (Aley *et al., J. Exp. Med.* (1984) 160:1585-1590; Leech *et al., J. Exp. Med.* (1984) 159:1567-1575; Howard *et al., Molec. Biochem. Parasitol.* (1988) 27:207-223), and expression of new receptor properties which mediate adherence to endothelial cells (Miller, supra; Udeinya *et al., Science* (1981) 213:555-557. Endothelial cell surface proteins such as CD36, thrombospondin (TSP) and ICAM-1 have been identified as major host

receptors for mature PE. *See, e.g.*, Barnwell *et al.*, *J. Immunol.* (1985) 135:3494-3497; Roberts *et al.*, *Nature* (1985) 318:64-66; and Berendt *et al.*, *Nature* (1989) 341:57-59.

PE sequestration confers unique advantages for P. falciparum parasites (Howard and Gilladoga, *Blood* (1989) 74:2603-2618), but also contributes directly to the acute pathology of P. falciparum (Miller *et al.*, *Science* (1994) 264:1878-1883). Of the four human malarias, only P. falciparum infection is associated with neurological impairment and cerebral pathology seen increasingly in severe drug-resistant malaria (Howard and Gilladoga, *supra*).

Although the genesis of human cerebral malaria is likely due to a combination of factors including particular parasite phenotypes (Berendt *et al.*, *Parasitol. Today* (1994) 10:412- 414), inappropriate immune responses and the phenotype of endothelial cell surface molecules in the cerebral microvasculature (Pasloske and Howard, *Ann. Rev. Med.* (1994) :283-295), adherence of PE to cerebral blood vessels and consequent local microvascular occlusion is a major contributing factor. *See, e.g.*, Berendt *et al.*, *supra*; Patnaik *et al.*, *Am. J. Trop. Med. Hyg.* (1994) 51:642-647.

The capacity of P. falciparum PE to express variant forms of PfEMP1 contributes to the special virulence of this parasite. Variant parasites can evade variant-specific antibodies elicited by earlier infections. The P. falciparum variant antigens have been defined *in vitro* using antiserum prepared in Aotus monkeys infected with individual parasite strains (Howard *et al.*, *Molec. Biochem. Parasitol.* (1988) 27:207-223). Antibodies raised against a particular parasite will only react by PE agglutination, indirect immuno- fluorescence or immuno-electron microscopy with PE from the same strain (van Schravendijk *et al.*, *Blood* (1991) 78:226-236).

Such studies with PE from malaria patients in diverse geographic locations and sera from the same or different patients confirm that PE in natural isolates express variant surface antigens and that individual patients respond to infection by production of isolate-specific antibodies (Marsh and Howard, *Science* (1986) 231:150-153; Aguiar et al., Am. J. Trop. Med. Hyg. (1992) 47:621-632; Iqbal et al., Trans. R. Soc. Trop. Med. Hyg. (1993) 87:583-588. Expression of a variant antigen on PE has also been demonstrated in several simian, murine and human malaria species, including *P. knowlesi* (Brown and Brown, Nature (1965) 208:1286-1288; Barnwell et al., Infect. Immun. (1983) 40:985-994), P. chabaudi (Gilks et al., Parasite Immunol. (1990) 12:45-64; Brannan et al., Proc. R. Soc. Lond. Biol. Sci. (1994) 256:71- 75), P. fragile (Handunnetti et al., J. Exp. Mod. (1987) 165:1269-1283) and P. vivax

(Mendis et al., Am. J. Txop. Med. Hyg. (1988) 38:42-46). Laboratory studies with P. knowlesi (Brown and Brown, supra; Barnwell et al., supra) or P. falciparum (Hommel et al., J. Exp. Med. (1983) 157:1137- 1148) in monkeys and P. chabaudi in mice (Gilks et al., supra) confirmed that antigenic variation at the PE surface is associated with prolonged or chronic infection and the capacity to repeatedly re-establish blood infection in previously infected animals. Studies with cloned parasites demonstrated that antigenic variants can arise with extraordinary frequency, e.g., 2% per generation with P. falciparum (Roberts et al., Nature (1992) 357:689-692) and 1.6 % per generation with P. chabaudi (Brannan et al., supra).

PfEMP1 was identified as a ^{125}I-labeled, size diverse protein (200-350 kD) on PE that is lacking from uninfected erythrocytes, and that is also labeled by biosynthetic incorporation of radiolabeled amino acids (Leech et al., J. Exp. Med. (1984) 159:1567-1575; Howard et al., Molec. Biochem. Parasitol. (1988) 27:207-223). PfEMP1 is not extracted from PE by neutral detergents such as Triton X-100 but is extracted by SDS, suggesting that it is linked to the erythrocyte cytoskeleton (Aley et al., J. Med. Exp. (1984) 160:1585-1590). After addition of excess Triton X-100, PfEMP1 is immunoreactive with appropriate serum antibodies (Howard et al., (1988), supra). Mild trypsinization of intact PE rapidly cleaves PfEMP1 from the cell surface (Leech et al., J. Exp. Mod. (1984) 159:1567-1575). PfEMP1 bears antigenically diverse epitopes since it is immunoprecipitated from particular strains of P. falciparum by antibodies from sera of Aotus monkeys infected with the same strain, but not by antibodies from animals infected with heterologous strains (Howard et al. (1988), supra). Knobless PE derived from parasite passage in splenectomized Aotus monkeys (Aley et al., supra) do not express surface PfEMP1 and are not agglutinated with sera from immune individuals or infected monkeys (Howard et al. (1988), supra; Howard and Gilladoga, Blood (1989) 74:2603-2618). In general, sera that react with the PE surface by indirect immunofluorescence and antibody-mediated PE agglutination are the only sera to immunoprecipitate ^{125}I-labeled PfEMP1 from any particular strain (Howard et al., (1988), supra; van Schravendijk et al., Blood (1991) 78:226- 236; Biggs et al., J. Immunol. (1992) 149:2047-2054).

The adherence of parasitized erythrocytes to endothelial cells is mediated by multiple receptor/counter- receptor interactions, including CD36, thrombospondin and intracellular adhesion molecule-1 (ICAM_1) as the major host cell receptors (Howard and Gilladoga, Blood (1989) 74:2603- 2618, Pasloske and Howard, Ann. Rev. Med. (1994) 45:283-295).

Vascular cell adhesion molecule-1 (VCAM-1) and endothelial leukocyte adhesion molecule-1 (ELAM-1) have also been implicated as additional endothelial cell receptors that can mediate adherence of a minority of P. falciparum PE (Ockenhouse, et al., J. Exp. Med. (1992) 176:1183-1189, and Howard and Paslaske, supra). The adherence receptors on the surface of PE has not yet been conclusively identified, and several molecules, including AG 332 (Udomsangpetch, et al., Nature (1989) 338:763-765), modified band 3 (Crandall, et al., Proc. Nat'l Acad. Sci. USA (1993) 90:4703-4707), Sequestrin (Ockenhouse, Proc. Nat'l Acad. Sci. USA (1991) 88:3175-3179), and PfEMP1 (Howard and Gilladoga, supra, and Pasloske and Howard, supra), have been proposed as candidates. Several pieces of indirect evidence have linked expression of PfEMP1 with the acquisition of new host protein receptor properties on the surface of PE (Howard and Gilladoga, supra; Pasloske and Howard, Ann. Rev. Med. (1994) 45:283-295). PE adherence is correlated with the expression of PfEMP1 on the surface of mature stage PE (Leech, et al., J. Exp. Med. (1984) 159:1567- 1575). Alterations in the adherence phenotype of the PE selected for *in vitro* are usually associated with the emergence of new forms of PfEMP1 (Biggs, et al., J. Immunol. (1992) 149:2047-2054; Roberts, et al., Nature (1992) 357:689- 692). Mild trypsinization of intact mature PE cleaves the extracellular portion of PfEMP1 and at the same time, reduces or eliminates PE cytoadherence (Leech, et al., supra) Previously described antibody mediated blockade or reversal of cytoadherence is strain specific and is correlated with the ability of the reacting sera to agglutinate the corresponding PE and to immunoprecipitate the surface labeled ^{125}I-PfEMP1 (Howard, et al., Molec. Biochem. Parasitol. (1988) 27:207- 224). Pfalhesin (modified band 3) have been shown to bind CD36 under non-physiological conditions (Crandall, et al., Exp. Parasitol. (1994) 78:203-209). Sequestrin, which appears to be homologous to PfEMP1, extracted with TX100 from knobless PE, was shown to bind to immobilized CD36 (Ockenhouse, Proc. Nat'l Acad. Sci. USA (1991) 88:3175-3179).

The complex nature and/or mechanism of malarial antigenic variation, and its particular virulence has created a need for methods and compositions which may be useful in the treatment diagnosis and prevention of malaria infections.

General Overview of Problems & Considerations in Directed Evolution

The approach, termed directed evolution, of experimentally modifying a biological molecule towards a desirable property, can be achieved by mutagenizing one or more parental molecular templates and by idendifying any desirable molecules among the progeny

molecules. Currently available technologies in directed evolution include methods for achieving stochastic (i.e. random) mutagenesis and methods for achieving non-stochastic (non-random) mutagenesis. However, critical shortfalls in both types of methods are identified in the instant disclosure.

In prelude, it is noteworthy that it may be argued philosophically by some that all mutagenesis – if considered from an objective point of view – is non-stochastic; and furthermore that the entire universe is undergoing a process that – if considered from an objective point of view – is non-stochastic. Whether this is true is outside of the scope of the instant consideration. Accordingly, as used herein, the terms "randomness", "uncertainty", and "unpredictability" have subjective meanings, and the knowledge, particularly the predictive knowledge, of the designer of an experimental process is a determinant of whether the process is stochastic or non-stochastic.

By way of illustration, stochastic or random mutagenesis is exemplified by a situation in which a progenitor molecular template is mutated (modified or changed) to yield a set of progeny molecules having mutation(s) that are not predetermined. Thus, in an in vitro stochastic mutagenesis reaction, for example, there is not a particular predetermined product whose production is intended; rather there is an uncertainty – hence randomness – regarding the exact nature of the mutations achieved, and thus also regarding the products generated. In contrast, non-stochastic or non-random mutagenesis is exemplified by a situation in which a progenitor molecular template is mutated (modified or changed) to yield a progeny molecule having one or more predetermined mutations. It is appreciated that the presence of background products in some quantity is a reality in many reactions where molecular processing occurs, and the presence of these background products does not detract from the non-stochastic nature of a mutagenesis process having a predetermined product.

Thus, as used herein, stochastic mutagenesis is manifested in processes such as error-prone PCR and stochastic shuffling, where the mutation(s) achieved are random or not predetermined. In contrast, as used herein, non-stochastic mutagenesis is manifested in instantly disclosed processes such as gene site-saturation mutagenesis and synthetic ligation reassembly, where the exact chemical structure(s) of the intended product(s) are predetermined.

In brief, existing mutagenesis methods that are non-stochastic have been serviceable in generating from one to only a very small number of predetermined mutations per method

application, and thus produce per method application from one to only a few progeny molecules that have predetermined molecular structures. Moreover, the types of mutations currently available by the application of these non-stochastic methods are also limited, and thus so are the types of progeny mutant molecules.

In contrast, existing methods for mutagenesis that are stochastic in nature have been serviceable for generating somewhat larger numbers of mutations per method application – though in a random fashion & usually with a large but unavoidable contingency of undesirable background products. Thus, these existing stochastic methods can produce per method application larger numbers of progeny molecules, but that have undetermined molecular structures. The types of mutations that can be achieved by application of these current stochastic methods are also limited, and thus so are the types of progeny mutant molecules.

There is a need for the development of non-stochastic mutagenesis methods that: 1) Can be used to generate large numbers of progeny molecules that have predetermined molecular structures; 2) Can be used to readily generate more types of mutations; 3) Can produce a correspondingly larger variety of progeny mutant molecules; 4) Produce decreased unwanted background products; 5) Can be used in a manner that is exhaustive of all possibilities; and 6) Can produce progeny molecules in a systematic & non-repetitive way.

<u>Directed Evolution Supplements Natural Evolution:</u>

Natural evolution has been a springboard for directed or experimental evolution, serving both as a reservoir of methods to be mimicked and of molecular templates to be mutagenized. It is appreciated that, despite its intrinsic process-related limitations (in the types of favored &/or allowed mutagenesis processes) and in its speed, natural evolution has had the advantage of having been in process for millions of years & and throughout a wide diversity of environments. Accordingly, natural evolution (molecular mutagenesis and selection in nature) has resulted in the generation of a wealth of biological compounds that have shown usefulness in certain commercial applications.

However, it is instantly appreciated that many unmet commercial needs are discordant with any evolutionary pressure &/or direction that can be found in nature. Moreover, it is often the case that when commercially useful mutations would otherwise be favored at the molecular level in nature, natural evolution often overrides the positive selection of such mutations, e.g.

when there is a concurrent detriment to an organism as a whole (such as when a favorable mutation is accompanied by a detrimental mutation). Additionally, natural evolution is often slow, and favors fidelity in many types of replication. Additionally still, natural evolution often favors a path paved mainly by consecutive beneficial mutations while tending to avoid a plurality of successive negative mutations, even though such negative mutations may prove beneficial when combined, or may lead - through a circuitous route - to final state that is beneficial.

Moreover, natural evolution advances through specific steps (e.g. specific mutagenesis and selection processes), with avoidance of less favored steps. For example, many nucleic acids do not reach close enough proximity to each other in a operative environment to undergo chimerization or incorporation or other types of transfers from one species to another. Thus, e.g., when sexual intercourse between 2 particular species is avoided in nature, the chimerization of nucleic acids from these 2 species is likewise unlikely, with parasites common to the two species serving as an example of a very slow passageway for inter-molecular encounters and exchanges of DNA. For another example, the generation of a molecule causing self-toxicity or self-lethality or sexual sterility is avoided in nature. For yet another example, the propagation of a molecule having no particular immediate benefit to an organism is prone to vanish in subsequent generations of the organism. Furthermore, e.g., there is no selection pressure for improving the performance of molecule under conditions other than those to which it is exposed in its endogenous environment; e.g. a cytoplasmic molecule is not likely to acquire functional features extending beyond what is required of it in the cytoplasm. Furthermore still, the propagation of a biological molecule is susceptible to any global detrimental effects – whether caused by itself or not – on its ecosystem. These and other characteristics greatly limit the types of mutations that can be propagated in nature.

On the other hand, directed (or experimental) evolution – particularly as provided herein – can be performed much more rapidly and can be directed in a more streamlined manner at evolving a predetermined molecular property that is commercially desirable where nature does not provide one &/or is not likely to provide. Moreover, the directed evolution invention provided herein can provide more wide-ranging possibilities in the types of steps that can be used in mutagenesis and selection processes. Accordingly, using templates harvested from nature, the instant directed evolution invention provides more wide-ranging possibilities in the

types of progeny molecules that can be generated and in the speed at which they can be generated than often nature itself might be expected to in the same length of time.

In a particular exemplification, the instantly disclosed directed evolution methods can be applied iteratively to produce a lineage of progeny molecules (e.g. comprising successive sets of progeny molecules) that would not likely be propagated (i.e., generated &/or selected for) in nature, but that could lead to the generation of a desirable downstream mutagenesis product that is not achievable by natural evolution.

Previous Directed Evolution Methods Are Suboptimal:

Mutagenesis has been attempted in the past on many occasions, but by methods that are inadequate for the purpose of this invention. For example, previously described non-stochastic methods have been serviceable in the generation of only very small sets of progeny molecules (comprised often of merely a solitary progeny molecule). By way of illustration, a chimeric gene has been made by joining 2 polynucleotide fragments using compatible sticky ends generated by restriction enzyme(s), where each fragment is derived from a separate progenitor (or parental) molecule. Another example might be the mutagenesis of a single codon position (i.e. to achieve a codon substitution, addition, or deletion) in a parental polynucleotide to generate a single progeny polynucleotide encoding for a single site-mutagenized polypeptide.

Previous non-stochastic approaches have only been serviceable in the generation of but one to a few mutations per method application. Thus, these previously described non-stochastic methods thus fail to address one of the central goals of this invention, namely the exhaustive and non-stochastic chimerization of nucleic acids. Accordingly previous non-stochastic methods leave untapped the vast majority of the possible point mutations, chimerizations, and combinations thereof, which may lead to the generation of highly desirable progeny molecules.

In contrast, stochastic methods have been used to achieve larger numbers of point mutations and/or chimerizations than non-stochastic methods; for this reason, stochastic methods have comprised the predominant approach for generating a set of progeny molecules that can be subjected to screening, and amongst which a desirable molecular species might hopefully be found. However, a major drawback of these approaches is that – because of their stochastic nature – there is a randomness to the exact components in each set of progeny molecules that is produced. Accordingly, the experimentalist typically has little or no idea

what exact progeny molecular species are represented in a particular reaction vessel prior to their generation. Thus, when a stochastic procedure is repeated (e.g. in a continuation of a search for a desirable progeny molecule), the re-generation and re-screening of previously discarded undesirable molecular species becomes a labor-intensive obstruction to progress, causing a circuitous – if not circular – path to be taken. The drawbacks of such a highly suboptimal path can be addressed by subjecting a stochastically generated set of progeny molecules to a labor-incurring process, such as sequencing, in order to identify their molecular structures, but even this is an incomplete remedy.

Moreover, current stochastic approaches are highly unsuitable for comprehensively or exhaustively generating all the molecular species within a particular grouping of mutations, for attributing functionality to specific structural groups in a template molecule (e.g. a specific single amino acid position or a sequence comprised of two or more amino acids positions), and for categorizing and comparing specific grouping of mutations. Accordingly, current stochastic approaches do not inherently enable the systematic elimination of unwanted mutagenesis results, and are, in sum, burdened by too many inherently shortcomings to be optimal for directed evolution.

An exceedingly large number of possibilities exist for the purposeful and random combination of amino acids within a protein to produce useful hybrid proteins and their corresponding biological molecules encoding for these hybrid proteins, i.e., DNA, RNA. Accordingly, there is a need to produce and screen a wide variety of such hybrid proteins for a desirable utility, particularly widely varying random proteins.

The complexity of an active sequence of a biological macromolecule (e.g., polynucleotides, polypeptides, and molecules that are comprised of both polynucleotide and polypeptide sequences) has been called its information content ("IC"), which has been defined as the resistance of the active protein to amino acid sequence variation (calculated from the minimum number of invariable amino acids (bits) required to describe a family of related sequences with the same function). Proteins that are more sensitive to random mutagenesis have a high information content.

Molecular biology developments, such as molecular libraries, have allowed the identification of quite a large number of variable bases, and even provide ways to select functional sequences from random libraries. In such libraries, most residues can be varied (although typically not all at the same time) depending on compensating changes in the

context. Thus, while a 100 amino acid protein can contain only 2,000 different mutations, 20^{100} sequence combinations are possible.

Information density is the IC per unit length of a sequence. Active sites of enzymes tend to have a high information density. By contrast, flexible linkers of information in enzymes have a low information density.

Current methods in widespread use for creating alternative proteins in a library format are error-prone polymerase chain reactions and cassette mutagenesis, in which the specific region to be optimized is replaced with a synthetically mutagenized oligonucleotide. In both cases, a substantial number of mutant sites are generated around certain sites in the original sequence.

Error-prone PCR uses low-fidelity polymerization conditions to introduce a low level of point mutations randomly over a long sequence. In a mixture of fragments of unknown sequence, error-prone PCR can be used to mutagenize the mixture. The published error-prone PCR protocols suffer from a low processivity of the polymerase. Therefore, the protocol is unable to result in the random mutagenesis of an average-sized gene. This inability limits the practical application of error-prone PCR. Some computer simulations have suggested that point mutagenesis alone may often be too gradual to allow the large-scale block changes that are required for continued and dramatic sequence evolution. Further, the published error-prone PCR protocols do not allow for amplification of DNA fragments greater than 0.5 to 1.0 kb, limiting their practical application. In addition, repeated cycles of error-prone PCR can lead to an accumulation of neutral mutations with undesired results, such as affecting a protein's immunogenicity but not its binding affinity.

In oligonucleotide-directed mutagenesis, a short sequence is replaced with a synthetically mutagenized oligonucleotide. This approach does not generate combinations of distant mutations and is thus not combinatorial. The limited library size relative to the vast sequence length means that many rounds of selection are unavoidable for protein optimization. Mutagenesis with synthetic oligonucleotides requires sequencing of individual clones after each selection round followed by grouping them into families, arbitrarily choosing a single family, and reducing it to a consensus motif. Such motif is re-synthesized and reinserted into a single gene followed by additional selection. This step process constitutes a statistical bottleneck, is labor intensive, and is not practical for many rounds of mutagenesis.

Error-prone PCR and oligonucleotide-directed mutagenesis are thus useful for single cycles of sequence fine tuning, but rapidly become too limiting when they are applied for multiple cycles.

Another limitation of error-prone PCR is that the rate of down-mutations grows with the information content of the sequence. As the information content, library size, and mutagenesis rate increase, the balance of down-mutations to up-mutations will statistically prevent the selection of further improvements (statistical ceiling).

In cassette mutagenesis, a sequence block of a single template is typically replaced by a (partially) randomized sequence. Therefore, the maximum information content that can be obtained is statistically limited by the number of random sequences (i.e., library size). This eliminates other sequence families which are not currently best, but which may have greater long term potential.

Also, mutagenesis with synthetic oligonucleotides requires sequencing of individual clones after each selection round. Thus, such an approach is tedious and impractical for many rounds of mutagenesis.

Thus, error-prone PCR and cassette mutagenesis are best suited, and have been widely used, for fine-tuning areas of comparatively low information content. One apparent exception is the selection of an RNA ligase ribozyme from a random library using many rounds of amplification by error-prone PCR and selection.

In nature, the evolution of most organisms occurs by natural selection and sexual reproduction. Sexual reproduction ensures mixing and combining of the genes in the offspring of the selected individuals. During meiosis, homologous chromosomes from the parents line up with one another and cross-over part way along their length, thus randomly swapping genetic material. Such swapping or shuffling of the DNA allows organisms to evolve more rapidly.

In recombination, because the inserted sequences were of proven utility in a homologous environment, the inserted sequences are likely to still have substantial information content once they are inserted into the new sequence.

Theoretically there are 2,000 different single mutants of a 100 amino acid protein. However, a protein of 100 amino acids has 20^{100} possible sequence combinations, a number which is too large to exhaustively explore by conventional methods. It would be

advantageous to develop a system which would allow generation and screening of all of these possible combination mutations.

Some workers in the art have utilized an *in vivo* site specific recombination system to generate hybrids of combine light chain antibody genes with heavy chain antibody genes for expression in a phage system. However, their system relies on specific sites of recombination and is limited accordingly. Simultaneous mutagenesis of antibody CDR regions in single chain antibodies (scFv) by overlapping extension and PCR have been reported.

Others have described a method for generating a large population of multiple hybrids using random *in vivo* recombination. This method requires the recombination of two different libraries of plasmids, each library having a different selectable marker. The method is limited to a finite number of recombinations equal to the number of selectable markers existing, and produces a concomitant linear increase in the number of marker genes linked to the selected sequence(s).

In vivo recombination between two homologous, but truncated, insect-toxin genes on a plasmid has been reported as a method of producing a hybrid gene. The *in vivo* recombination of substantially mismatched DNA sequences in a host cell having defective mismatch repair enzymes, resulting in hybrid molecule formation has been reported.

SUMMARY

The invention provides a method for producing a plurality of, or a library of, nucleic acids encoding a plurality of modified antigen binding sites, wherein the modified antigen binding sites are derived from a first nucleic acid comprising a sequence encoding a first antigen binding site, the method comprising: (a) providing a first nucleic acid encoding a first antigen binding site; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of antigen binding site-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of modified antigen binding sites.

In one aspect of the method of the invention, step (b) provides a set of mutagenic oligonucleotides that encode all nineteen naturally-occurring amino acid variants for each

targeted codon, thereby generating all 19 possible natural amino acid changes at each amino acid codon mutagenized.

The method can further comprise expressing the set of variant antigen binding site-encoding nucleic acids such that antigen binding site-encoding polypeptides encoded by the variant nucleic acids are expressed.

In one aspect, the set of mutagenic oligonucleotides comprises a 19-fold degenerate mutagenic oligonucleotide for each codon to be mutagenized, wherein each of the 19-fold degenerate mutagenic oligonucleotides comprises a homologous first sequence and a degenerate triplet second sequence.

The antigen binding site can comprise a single stranded antigen binding polypeptide, a Fab fragment, an Fc fragment, a Fd fragment, a $F(ab')_2$ fragment, a Fv fragment or a complementarity determining region (CDR). The antigen binding site polypeptide can further comprise an antibody polypeptide.

In another aspect, the antigen binding site polypeptide further comprises an antigen binding site of a T cell receptor (TCR). The TCR antigen binding site polypeptide sequence modified by the methods of the invention can include the TCR alpha chain, the TCR beta chain, or both. The antigen binding site polypeptide can further comprise a T cell receptor (TCR).

In another aspect, the antigen binding site polypeptide further comprises an antigen binding site of a major histocompatibility complex (MHC) molecule. The antigen binding site polypeptide can further comprise a major histocompatibility complex (MHC) molecule. In alternative aspects, the major histocompatibility complex (MHC) molecule can comprise a Class I MHC molecule or a Class II MHC molecule. The MHC antigen binding site polypeptide sequence modified by the methods of the invention can include the MHC Class II alpha chain, the MHC Class II beta chain, or both.

In alternative aspects, the nucleic acid of step (a) is derived from a nucleic acid encoding a mammalian polypeptide, such as a human polypeptide. The mammalian polypeptide can be an antibody, a T cell receptor (alpha chain and/or beta chain), a Class I MHC molecule or a Class II MHC molecule(alpha chain and/or beta chain).

The nucleic acid of step (a) can be derived from a human nucleic acid encoding an antigen binding site. The nucleic acid of step (a) can be derived from a phage comprising a human nucleic acid sequence encoding an antigen binding site, wherein the phage expresses

the antigen binding site. The nucleic acid of step (a) can be derived from a non-human mammal comprising a human nucleic acid sequence encoding an antigen binding site, wherein the non-human mammal expresses the antigen binding site. The non-human mammal can be a transgenic non-human mammal, such as a mouse.

In one aspect of the method, at least two amino acid codons in the antigen binding site are mutagenized. Alternatively, all the amino acid codons in the antigen binding site are mutagenized, or, all the amino acid codons in the protein, e.g., the antibody, T cell receptor (TCR) or MHC polypeptide are mutagenized.

In one aspect, the degenerate mutagenic oligonucleotide comprises a first homologous sequence, a degenerate triplet second sequence, and a third homologous sequence. In another aspect, each degenerate oligonucleotide comprises a first homologous sequence, a plurality of degenerate triplets second sequences, and a third homologous sequence.

The method can further comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen. In one aspect, the method further comprises screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen capable of being specifically bound by the first antigen binding site polypeptide. In one aspect, the method comprises identifying an antigen binding site variant by its increased antigen binding affinity or antigen binding specificity as compared to the affinity or specificity of the first antigen binding site to the antigen. In one aspect, the method comprises identifying an antigen binding site variant by its decreased antigen binding affinity or antigen binding specificity as compared to the affinity or specificity of the first antigen binding site to the antigen.

The method can comprise mutagenizing the first nucleic acid of step (a) by a method comprising an optimized directed evolution system. The method can comprise mutagenizing the first nucleic acid of step (a) by a method comprising a synthetic ligation reassembly.

The method can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising expression of the expressed antigen binding site polypeptide in a solid phase. The method can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising a capillary array.

The method can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising a double-orificed container, such as a double-orificed capillary array, e.g., a GIGAMATRIX™ capillary array. The method can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising use of an ELISA. The method also can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising phage display of the antigen binding site polypeptide. The method also can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising expression of the expressed antigen binding site polypeptide in a liquid phase. The method also can comprise screening the expressed antigen binding site polypeptide for its ability to specifically bind an antigen by a method comprising ribosome display of the antigen binding site polypeptide.

In one aspect of the method, the set of progeny antigen binding site-encoding variant nucleic acids is generated by amplifying the nucleic acid of step (a) by a polymerase-based amplification using a plurality of oligonucleotides, such as polymerase chain reaction (PCR).

The invention provides a library of nucleic acids encoding a plurality of modified antigen binding sites, wherein the modified antigen binding sites are derived from a first nucleic acid comprising a sequence encoding a first antigen binding site, made by a method comprising the following steps: (a) providing a first nucleic acid encoding a first antigen binding site; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of antigen binding site-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of modified antigen binding sites.

The invention provides a method for producing from a library of variant antibodies from a template antibody, the method comprising: (a) providing a first nucleic acid encoding the template antibody; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, c) using the set of mutagenic oligonucleotides to generate a set of antibody-encoding variant nucleic acids encoding a range of amino acid variations at each

amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of variant antibodies. In one aspect, of the method, step (b) provides a set of mutagenic oligonucleotides that encode all nineteen naturally-occurring amino acid variants for each targeted codon, thereby generating all 19 possible natural amino acid changes at each amino acid codon mutagenized. The antibody can be a polypeptide comprising a Fab fragment, an Fd fragment, an Fc fragment, a F(ab')2 fragment, a Fv fragment or a complementarity determining region (CDR).

In one aspect of the method, the plurality of oligonucleotides comprises a degenerate oligonucleotide for each codon to be mutagenized, wherein each of the degenerate oligonucleotides comprises a homologous first sequence and a degenerate triplet second sequence. The set of progeny polynucleotides encoding antibodies can be generated by amplifying the nucleic acid of step (a) using a plurality of oligonucleotides.

The invention provides a library of variant antibodies derived from a template antibody made by a method comprising the following steps: (a) providing a first nucleic acid encoding the template antibody; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of antibody-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of variant antibodies.

The invention provides a method for producing from a library of variant T cell receptors (TCRs) from a template T cell receptor (TCR), the method comprising: (a) providing a first nucleic acid encoding the template T cell receptor; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, c) using the set of mutagenic oligonucleotides to generate a set of T cell receptor (TCR)-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of variant T cell receptors (TCRs).

The invention provides a library of variant T cell receptors (TCRs) derived from a template T cell receptor (TCR) made by a method comprising the following steps: (a) providing a first nucleic acid encoding the template T cell receptor; (b) providing a set of

mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of T cell receptor (TCR)-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of variant T cell receptors (TCRs).

The invention provides a method for producing from a library of variant major histocompatibility complex (MHC) molecules from a template major histocompatibility complex (MHC) molecule, the method comprising: (a) providing a first nucleic acid encoding the template major histocompatibility complex (MHC) molecule; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of major histocompatibility complex (MHC) molecule-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of variant major histocompatibility complex (MHC) molecules.

The invention provides a library of variant major histocompatibility complex (MHC) molecules derived from a template major histocompatibility complex (MHC) molecule made by a method comprising the following steps: (a) providing a first nucleic acid encoding the template major histocompatibility complex (MHC) molecule; (b) providing a set of mutagenic oligonucleotides that encode naturally-occurring amino acid variants at a plurality of targeted codons in the first nucleic acid; and, (c) using the set of mutagenic oligonucleotides to generate a set of major histocompatibility complex (MHC) molecule-encoding variant nucleic acids encoding a range of amino acid variations at each amino acid codon that was mutagenized, thereby producing a library of nucleic acids encoding a plurality of variant major histocompatibility complex (MHC) molecules.

The invention provides a method of making a set of nucleic acids encoding a set of antigen binding site variants comprising the steps of: (a) providing a template nucleic acid encoding an antigen-binding polypeptide; (b) providing a plurality of oligonucleotides that encode all nineteen naturally-occurring amino acid variants at a single amino acid residue of the antigen-binding polypeptide; and, (c) generating a set of progeny antigen binding site-encoding variant nucleic acids encoding a non-stochastic range of single amino acid

substitutions at each amino acid codon that was mutagenized, whereby all 19 possible natural amino acid changes are generated at each amino acid codon mutagenized, thereby making a set of nucleic acids encoding a set of antigen binding site variants. In one aspect of the invention, the method further comprises expressing the set of progeny antigen binding site-encoding polynucleotides such that antigen binding site-encoding polypeptides encoded by the progeny polynucleotides are expressed. The plurality of oligonucleotides can comprise a set of degenerate oligonucleotides and each of the degenerate oligonucleotides comprises a homologous first sequence and a degenerate triplet second sequence.

In one aspect, the antigen binding site-encoding polypeptide comprises a single stranded antigen binding polypeptide. The antigen binding site-encoding polypeptide can comprise an antibody polypeptide. The antigen binding site-encoding polypeptide can comprise an antigen binding site of a T cell receptor (TCR), or, a T cell receptor (TCR). In alternative aspects, the antigen binding site-encoding polypeptide can comprise an antigen binding site of a major histocompatibility complex (MHC) molecule, or, a major histocompatibility complex (MHC) molecule.

In one aspect, the nucleic acid of step (a) can be derived from a nucleic acid encoding a mammalian antibody polypeptide. The nucleic acid of step (a) can be derived from a human nucleic acid.

In one aspect, the at least two amino acid codons in the antigen binding site are mutagenized and a set of degenerate oligonucleotides that encode all nineteen naturally-occurring amino acid variants are provided for each amino acid codon mutagenized. In one aspect, all the amino acid codons in the antigen binding site are mutagenized and a set of degenerate oligonucleotides that encode all nineteen naturally-occurring amino acid variants are provided for each amino acid codon mutagenized.

In one aspect, all the amino acid codons in the antibody polypeptide can be mutagenized. In alternative aspects, all the amino acid codons in the antigen binding site of the T cell receptor (TCR) are mutagenized, all the amino acid codons in the antigen binding site of the major histocompatibility complex (MHC) molecule are mutagenized; and all the amino acid codons in the antigen binding site of the antibody are mutagenized.

In one aspect, a degenerate oligonucleotide comprises a first homologous sequence, a degenerate triplet second sequence, and a homologous third sequence. In one aspect, each

degenerate oligonucleotide comprises a first homologous sequence, a degenerate triplet second sequence, and a homologous third sequence.

In alternative aspects, the method further comprises mutagenizing the template nucleic acid by a method comprising an optimized directed evolution system and a method comprising a synthetic ligation reassembly.

In one aspect, the method further comprises screening an expressed antigen binding site-encoding polypeptide for its ability to specifically bind an antigen. The method can also comprise screening the expressed antigen binding site-encoding polypeptide for its ability to specifically bind an antigen capable of being specifically bound by the first antigen binding site. The method can comprise identifying an antigen binding site variant by its increased or decreased or altered antigen binding affinity or antigen binding specificity to the antigen as compared to the affinity or specificity of the antigen binding site encoded by the nucleic acid of step (a).

In alternative aspects, the method comprises screening the expressed antigen binding site-encoding polypeptide for its ability to specifically bind an antigen in a solid phase or a liquid phase. In one aspect, the method comprises a capillary array, such as a double-orificed capillary array. The method can comprise screening the expressed antigen binding site-encoding polypeptide for its ability to specifically bind an antigen by an ELISA.

In alternative aspects, the set of variant nucleic acids is generated by performing amplification reactions on the nucleic acid of step (a) using the set of oligonucleotides to generate a set of variant nucleic acids encoding nineteen amino acid substitution variants at least one amino acid residue of the antigen-binding polypeptide, or, all of the amino acid residue of the antigen-binding polypeptide. The amplification can comprise a polymerase-based amplification, such as a polymerase chain reaction (PCR), or another equivalent reaction.

In alternative aspects, the set of variant nucleic acids comprises 10^{10} members, 10^9, 10^8, 10^7, 10^6 10^5, 10^4, 10^3, 10^2 members.

The invention provides a method of making a set (i.e., a library) of antibody variants comprising the steps of: (a) providing a nucleic acid encoding an antibody; (b) providing a plurality of oligonucleotides; (c) generating a non-stochastic range of single amino acid substitutions at each amino acid codon, whereby all 19 possible natural amino acid changes are generated at each amino acid codon mutagenized, thereby generating a set of variant

nucleic acids; and, (d) expressing the set of variant nucleic acids such that the antibody variants encoded by the variant nucleic acids are expressed. The antibody can be selected from the group consisting of polypeptides comprising a Fab fragment, a Fd fragment, an Fc fragment, a $F(ab')_2$ fragment, a Fv fragment and a complementarity determining region (CDR). The plurality of oligonucleotides can comprise a set of degenerate oligonucleotides that encode all nineteen naturally-occurring amino acid variants at a single amino acid residue of the antibody, wherein each of the degenerate oligonucleotides comprises a homologous first sequence and a degenerate triplet second sequence. The method, in generating a non-stochastic range of single amino acid substitutions, can comprise performing amplification reactions on the nucleic acid of step (a) using the set of oligonucleotides to generate a set of variant nucleic acids encoding nineteen amino acid substitution variants at a single amino acid residue of the antibody.

The invention provides a method of identifying a variant of an antigen binding site comprising the steps of: (a) providing a nucleic acid encoding an antigen binding site; (b) providing a set of oligonucleotides that encode all nineteen naturally- occurring amino acid variants at all residues of the antigen- binding site; (c) incorporating the sequence of the oligonucleotides of step (b) into the nucleic acid of step (a) to generate a set of variant nucleic acids encoding nineteen amino acid substitution variants at each residue of the antigen binding site; (d) expressing each of the variant nucleic acids as polypeptides and measuring the variant's affinity to the antigen; and, (e) identifying a variant of the antigen binding site by its increased or decreased antigen binding specificity as compared to the antigen binding affinity of the antigen binding site encoded by the nucleic acid of step (a). In one aspect, the variant nucleic acids are expressed using *in vitro* transcription/translation. In alternative aspects, the variant nucleic acids are expressed using phage display, ribosome display, or equivalent methods. In alternative aspects, the method comprises screening the expressed antigen binding site for its ability to specifically bind an antigen in a solid phase or a liquid phase. In one aspect, the screening is accomplished using a double orificed container, such as a using a double orificed capillary array.

In one aspect, the set of oligonucleotides comprises a set of degenerate oligonucleotides that encode all nineteen naturally-occurring amino acid variants at at least one amino acid residue of the antibody, wherein each of the degenerate oligonucleotides comprises a homologous first sequence and a degenerate triplet second sequence. In one

aspect, the set of oligonucleotides comprises a set of degenerate oligonucleotides that encode all nineteen naturally-occurring amino acid variants at all amino acid residues of the antibody. In one aspect, the method incorporates the sequence of the oligonucleotides of step (b) into the nucleic acid of step (a) is accomplished by an amplification reaction using the oligonucleotides as primers.

In one aspect, the antigen binding site comprises an antibody, including a Fab fragment, an Fd fragment, an Fc fragment, a $F(ab')_2$ fragment, a Fv fragment and a complementarity determining region (CDR). In alternative aspects, the antigen binding site comprises an antigen binding site of a T cell receptor and a major histocompatibility complex molecule.

In alternative aspects, the antigen binding site-encoding nucleic acids generated by the methods of the invention (e.g., the libraries of nucleic acids encoding modified antigen binding sites) are further changed or "evolved." These nucleic acid sequences can be changed by mutagenesis, base residue insertion(s) or base residue deletion(s). Evolution technologies can be used to further engineer these sequences, including, e.g., Gene Site Saturation MutagenesisTM (GSSM) and GeneReassemblyTM (Diversa Corporation, San Diego, CA), as described in further detail herein. Alternatively, these nucleic acid sequences can be changed or "evolved" or "genetically engineered" by, e.g., error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturated mutagenesis (GSSM), synthetic ligation reassembly (SLR) and/or a combination thereof. In alternative aspects, the modifications, additions or deletions are introduced by, e.g., recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and/or a combination thereof. In one aspect, these methods are iteratively repeated until an antigen binding site (e.g., an antibody) having an altered or different activity or an altered or different stability from that of the antigen binding site to be

"evolved" is produced. In one aspect, the CDR3 region of the antigen binding molecule-encoding nucleic acid sequence is changed or "evolved."

In a non-limiting aspect, the instant invention provides non-stochastic means for comprehensively and exhaustively generating all possible point mutations in a parental template. In another non-limiting aspect, the instant invention further provides means for exhaustively generating all possible chimerizations within a group of chimerizations. Thus, the aforementioned problems (in background) are solved by the instant invention.

Specific shortfalls in the technological landscape addressed by this invention include, e.g., 1) Site-directed mutagenesis technologies, such as sloppy or low-fidelity PCR, are ineffective for systematically achieving at each position (site) along a polypeptide sequence the full (saturated) range of possible mutations (i.e. all possible amino acid substitutions). 2) There is no relatively easy systematic means for rapidly analyzing the large amount of information that can be contained in a molecular sequence and in the potentially colossal number or progeny molecules that could be conceivably obtained by the directed evolution of one or more molecular templates. 3) There is no relatively easy systematic means for providing comprehensive empirical information relating structure to function for molecular positions. 4) There is no easy systematic means for incorporating internal controls, such as positive controls, for key steps in certain mutagenesis (e.g. chimerization) procedures. 5) There is no easy systematic means to select for a specific group of progeny molecules, such as full-length chimeras, from among smaller partial sequences.

<u>Directing an immune response so as to achieve an optimal response to vaccination.</u>

The present invention provides multicomponent genetic vaccines that include at least one, or, two or more, genetic vaccine components that confer upon the vaccine the ability to direct an immune response so as to achieve an optimal response to vaccination. For example, the genetic vaccines can include a component that provides optimal antigen release; a component that provides optimal production of cytotoxic T lymphocytes; a component that directs release of an immunomodulator; a component that directs release of a chemokine; and/or a component that facilitates binding to, or entry into, a desired target cell type. For example, a component can confer improved improves binding to, and uptake of, the genetic vaccine to target cells such as antigen- expressing cells or antigen-presenting cells.

Additional components include those that direct antigen peptides derived from uptake of an antigen into a cell to presentation on either Class I or Class II molecules. For example,

one can include a component that directs antigen peptides to presentation on Class I molecules and comprises a polynucleotide that encodes a protein such as tapasin, TAP-1 and TAP-2, and/or a component that directs antigen peptides to presentation on Class II molecules and comprises a polynucleotide that encodes a protein such as an endosomal or lysosomal protease.

In one aspect, this invention provides a method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising: creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; wherein optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation and iterative manner; whereby these directed evolution methods include the introduction of mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein; and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

In one aspect, this invention provides a method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising: screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has, or encodes a polypeptide that has, a modulatory effect on an immune response; wherein the optimized non-stochastically generated polynucleotide or the polypeptide encoded by the non-stochastically generated polynucleotide exhibits an enhanced ability to modulate an immune response compared to a parental polynucleotide from which the library was created.

In one aspect, this invention provides a method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising: a) creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; and b) screening the library to identify an optimized non-stochastically generated progeny polynucleotide that has, or encodes a polypeptide that has, a modulatory effect on an immune response induced by a genetic vaccine vector; wherein the

optimized non-stochastically generated polynucleotide or the polypeptide encoded by the non-stochastically generated polynucleotide exhibits an enhanced ability to modulate an immune response compared to a parental polynucleotide from which the library was created; whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner; whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein; and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

In one aspect, this invention provides a method for obtaining an immunomodulatory polynucleotide that has, an optimized expression in a recombinant expression host, the method comprising: creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation and iterative manner; whereby these directed evolution methods include the introduction of mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein; and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

In one aspect, this invention provides a method for obtaining an immunomodulatory polynucleotide that has an optimized expression in a recombinant expression host, the method comprising: screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has an optimized expression in a recombinant expression host when compared to the expression of a parental polynucleotide from which the library was created.

In one aspect, this invention provides a method for obtaining an immunomodulatory polynucleotide that has an optimized expression in a recombinant expression host, the method comprising: a) creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; and b) screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically

generated progeny polynucleotide that has an optimized expression in a recombinant expression host when compared to the expression of a parental polynucleotide from which the library was created; whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner; whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein; and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

In one aspect, this invention provides that the ability to a vaccine, for example a genetic vaccine, or a component of a vaccine, for example a component of a genetic vaccine by optimizing its immunogenicity. Moreover, the present invention provides for the modification of other properties, including its:

- Catalyzed reaction(s)
- Reaction type
- Natural substrate(s)
- Substrate spectrum
- Product spectrum
- Inhibitor(s)
- Cofactor(s)/prosthetic group(s)
- Metal compounds/salts that affect it
- Turnover number
- Specific activity
- Km value
- pH optimum
- pH range
- Temperature optimum
- Temperature range

It is also instantly appreciated that the serviceability of a molecule with an immunogenic effect can be affected by additional physical properties, which can likewise be modified by directed evolution as provided herein, such as how it is affected by subjection to:

- Isolation/Preparation
- Purification
- Renaturation conditions (reversibility or retention of activity upon: heating and cooling, urea, salts, detergents, pH extremes)
- Crystallization
- pH

- Temperature
- Oxidation
- Organic solvent(s)
- Miscellaneous storage conditions

Moreover, the instant invention provides for the modification of molecule's immunogenic properties such as

- Exposure to biological compartments (stomach acids, in vivo degradation)
- Expression (e.g. Transcription &/or Translation) level
- mRNA stability
- Any *in vivo* interactions with other cells or biologicals

Method for obtaining the genetic components

In some embodiments, one or more of the genetic vaccine components is obtained by a method that involves: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which can confer a desired property upon a genetic vaccine, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids; and (2) screening the library to identify at least one optimized recombinant component that exhibits an enhanced capacity to confer the desired property upon the genetic vaccine. If further optimization of the component is desired, the following additional steps can be conducted: (3) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least one optimized recombinant component with a further form of the nucleic acid, which is the same or different from the first and second forms, to produce a further library of recombinant nucleic acids; (4) screening the further library to identify at least one further optimized recombinant component that exhibits an enhanced capacity to confer the desired property upon the genetic vaccine; and (5) repeating (3) and (4), as necessary, until the further optimized recombinant component exhibits a further enhanced capacity to confer the desired property upon the genetic vaccine.

Members of a gene family

In some aspects of the invention, the first form of the nucleic acid is a first member of a gene family and the second form of the nucleic acid comprises a second member of the gene family. Additional forms of the module nucleic acid can also be members of the gene family. As an example, the first member of the gene family can be obtained from a first species of organism and the second member of the gene family obtained from a second

species of organism. If desired, the optimized recombinant genetic vaccine component obtained by the methods of the invention can be backcrossed by, for example, reassembling (&/or subjecting to one or more directed evolution methods described herein) the optimized recombinant genetic vaccine component with a molar excess of one or both of the first and second forms of the substrate nucleic acids to produce a further library of recombinant genetic vaccine components; and screening the further library to identify at least one optimized recombinant genetic vaccine component that further enhances the capability of a genetic vaccine vector that includes the component to modulate the immune response.

Methods of obtaining a genetic vaccine component that confers upon a genetic vaccine vector an enhanced ability to replicate in a host cell.

Additional embodiments of the invention provide methods of obtaining a genetic vaccine component that confers upon a genetic vaccine vector an enhanced ability to replicate in a host cell. These methods involve creating a library of recombinant nucleic acids by subjecting to reassembly (&/or one or more additional directed evolution methods described herein) at least two forms of a polynucleotide that can confer episomal replication upon a vector that contains the polynucleotide; introducing into a population of host cells a library of vectors, each of which contains a member of the library of recombinant nucleic acids and a polynucleotide that encodes a cell surface antigen; propagating the population of host cells for multiple generations; and identifying cells which display the cell surface antigen on a surface of the cell, wherein cells which display the cell surface antigen are likely to harbor a vector that contains a recombinant vector module which enhances the ability of the vector to replicate episomally.

Obtaining genetic vaccine components that confer upon a vector an enhanced ability to replicate in a host cell.

Genetic vaccine components that confer upon a vector an enhanced ability to replicate in a host cell can also be obtained by creating a library of recombinant nucleic acids by subjecting to reassembly (&/or one or more additional directed evolution methods described herein) at least two forms of a polynucleotide derived from a human papillomavirus that can confer episomal replication upon a vector that contains the polynucleotide; introducing a library of vectors, each of which contains a member of the library of recombinant nucleic acids, into a population of host cells; propagating the host cells for a plurality of generations; and identifying cells that contain the vector.

In additional embodiments, the invention provides methods obtaining a genetic vaccine component that confers upon a vector an enhanced ability to replicate in a human host cell by creating a library of recombinant nucleic acids by subjecting to reassembly (&/or one or more additional directed evolution methods described herein) at least two forms of a polynucleotide that can confer episomal replication upon a vector that contains the polynucleotide; introducing a library of genetic vaccine vectors, each of which comprises a member of the library of recombinant nucleic acids, into a test system that mimics a human immune response; and determining whether the genetic vaccine vector replicates or induces an immune response in the test system. A suitable test system can involve human skin cells present as a xenotransplant on skin of an immunocompromised non-human host animal, for example, or a non-human mammal that comprises a functional human immune system. Replication in these systems can be detected by determining whether the animal exhibits an immune response against the antigen.

The invention also provides methods of obtaining a genetic vaccine component that confers upon a genetic vaccine an enhanced ability to enter an antigen- presenting cell. These methods involve creating a library of recombinant nucleic acids by subjecting to reassembly (&/or one or more additional directed evolution methods described herein) at least two forms of a polynucleotide that can confer episomal replication upon a vector that contains the polynucleotide; introducing a library of genetic vaccine vectors, each of which comprises a member of the library of recombinant nucleic acids, into a population of antigen-presenting or antigen-processing cells; and determining the percentage of cells in the population which contain the nucleic acid vector. Antigen- presenting or antigen-processing cells of interest include, for example, B cells, monocytes/macrophages, dendritic cells, Langerhans cells, keratinocytes, and muscle cells.

The present invention provides methods of obtaining a polynucleotide that has a modulatory effect on an immune response, including a T cell receptors, major histocompatibility complex (MHC) molecules, antibodies, or those induced by a genetic vaccine, either directly (i.e., as an immunomodulatory polynucleotide) or indirectly (i.e., upon translation of the polynucleotide to create an immunomodulatory polypeptide. The methods of the invention involve: creating a library of experimentally generated (in vitro &/or in vivo) polynucleotides; and screening the library to identify at least one optimized experimentally generated (in vitro &/or in vivo) polynucleotide that exhibits, either by itself or through the encoded polypeptide, an enhanced ability to modulate an immune response than a form of the nucleic acid from which the library was created. Examples include, for example, CpG-rich polynucleotide sequences, polynucleotide sequences that encode a costimulator (e.g., B7-1, B7-2, CD1, CD40, CD154 (ligand for CD40), CD150 (SLAM), or a cytokine. The screening step used in these methods can include, for example, introducing genetic vaccine vectors which comprise the library of recombinant nucleic acids into a cell, and identifying cells which exhibit an increased ability to modulate an immune response of interest or increased ability to express an immunomodulatory molecule. For example, a library of recombinant cytokine-encoding nucleic acids can be screened by testing the ability of cytokines encoded by the nucleic acids to activate cells which contain a receptor for the cytokine. The receptor for the cytokine can be native to the cell, or can be expressed from a heterologous nucleic acid that encodes the cytokine receptor. For example, the optimized costimulators can be tested to identify those for which the cells or culture medium are capable of inducing a predominantly T_H2 immune response, or a predominantly T_H1 immune response.

In some embodiments, the polynucleotide that has a modulatory effect on an immune response is obtained by: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid that is, or encodes a molecule that is, involved in modulating an immune response, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of experimentally generated (in vitro &/or in vivo) polynucleotides; and (2) screening the library to identify at least one optimized experimentally generated (in vitro &/or in vivo) polynucleotide that exhibits, either by itself or through the encoded polypeptide, an enhanced ability to modulate an immune response than a form of the nucleic acid from which the

library was created. If additional optimization is desired, the method can further involve: (3) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least one optimized experimentally generated (in vitro &/or in vivo) polynucleotide with a further form of the nucleic acid, which is the same or different from the first and second forms, to produce a further library of experimentally generated (in vitro &/or in vivo) polynucleotides; (4) screening, the further library to identify at least one further optimized experimentally generated (in vitro &/or in vivo) polynucleotide that exhibits an enhanced ability to modulate an immune response than a form of the nucleic acid from which the library was created.; and (5) repeating (3) and (4), as necessary, until the further optimized experimentally generated (in vitro &/or in vivo) polynucleotide exhibits an further enhanced ability to modulate an immune response than a form of the nucleic acid from which the library was created.

In some embodiments of the invention, the library of experimentally generated (in vitro &/or in vivo) polynucleotides is screened by: expressing the experimentally generated (in vitro &/or in vivo) polynucleotides so that the encoded peptides or polypeptides are produced as fusions with a protein displayed on the surface of a replicable genetic package; contacting the replicable genetic packages with a plurality of cells that display the receptor; and identifying cells that exhibit a modulation of an immune response mediated by the receptor.

The invention also provides methods for obtaining a polynucleotide that encodes an accessory molecule that improves the transport or presentation of antigens by a cell. These methods involve creating a library of experimentally generated (in vitro &/or in vivo) polynucleotides by subjecting to reassembly (&/or one or more additional directed evolution methods described herein) nucleic acids that encode all or part of the accessory molecule; and screening the library to identify an optimized experimentally generated (in vitro &/or in vivo) polynucleotide that encodes a recombinant accessory molecule that confers upon a cell an increased or decreased ability to transport or present an antigen on a surface of the cell compared to an accessory molecule encoded by the non-recombinant nucleic acids. In some embodiments, the screening step involves: introducing the library of experimentally generated (in vitro &/or in vivo) polynucleotides into a genetic vaccine vector that encodes an antigen to form a library of vectors; introducing the library of vectors into mammalian

cells; and identifying mammalian cells that exhibit increased or decreased immunogenicity to the antigen.

In some embodiments of the invention, the cytokine that is optimized is interleukin-12 and the screening is performed by growing mammalian cells which contain the genetic vaccine vector in a culture medium, and detecting whether T cell proliferation or T cell differentiation is induced by contact with the culture medium. In another embodiment, the cytokine is interferon-α and the screening is performed by expressing the recombinant vector module as a fusion protein which is displayed on the surface of a bacteriophage to form a phage display library, and identifying phage library members which are capable of inhibiting proliferation of a B cell line. Another embodiment utilizes B7-1 (CD80) or B7-2 (CD86) as the costimulator and the cell or culture medium is tested for ability to modulate an immune response.

The invention provides methods of using stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly to obtain optimized recombinant vector modules that encode cytokines and other costimulators that exhibit reduced immunogenicity compared to a corresponding polypeptide encoded by a non-optimized vector module. The reduced immunogenicity can be detected by introducing a cytokine or costimulator encoded by the recombinant vector module into a mammal and determining whether an immune response is induced against the cytokine.

The invention also provides methods of obtaining optimized immunomodulatory sequences that encode a cytokine antagonist. For example, suitable cytokine agonists include a soluble cytokine receptor and a transmembrane cytokine receptor having, a defective signal sequence. Examples include sIL-10R and sIL- 4R, and the like.

The present invention provides methods for obtaining a cell-specific binding molecule that is useful for increasing uptake or specificity of a genetic vaccine to a target cell. The methods involve: creating a library of experimentally generated (in vitro &/or in vivo) polynucleotides that by reassembling (&/or subjecting to one or more directed evolution methods described herein) a nucleic acid that encodes a polypeptide that comprises a nucleic acid binding domain and a nucleic acid that encodes a polypeptide that comprises a cell-specific binding domain; and screening the library to identify a experimentally generated (in vitro &/or in vivo) polynucleotide that encodes a binding molecule that can bind to a nucleic acid and to a cell-specific receptor. Target cells of particular interest include antigen-

presenting and antigen-processing cells, such as muscle cells, monocytes, dendritic cells, B cells, Langerhans cells, keratinocytes, and M-cells.

In some embodiments, the methods of the invention for obtaining a cell- specific binding moiety useful for increasing uptake or specificity of a genetic vaccine to a target cell involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprises a polynucleotide that encodes a nucleic acid binding domain and at least first and second forms of a nucleic acid which comprises a cell-specific ligand that specifically binds to a protein on the surface of a cell of interest, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant binding moiety- encoding nucleic acids; (2) transfecting into a population of host cells a library of vectors, each of which comprises: a) a binding site specific for the nucleic acid binding domain and b) a member of the library of recombinant binding moiety-encoding nucleic acids, wherein the recombinant binding moiety is expressed and binds to the binding site to form a vector-binding moiety complex; (3) lysing the host cells under conditions that do not disrupt binding of the vector-binding moiety complex; (4) contacting the vector- binding moiety complex with a target cell of interest; and (5) identifying target cells that contain a vector and isolating the optimized recombinant cell-specific binding moiety nucleic acids from these target cells. If further optimization is desired, the methods can further involve: (6) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least one optimized recombinant binding moiety-encoding nucleic acid with a further form of the polynucleotide that encodes a nucleic acid binding domain and/or a further form of the polynucleotide that encodes a cell-specific ligand, which are the same or different from the first and second forms, to produce a further library of recombinant binding moiety-encoding nucleic acids; (7) transfecting into a population of host cells a library of vectors that comprise: a) a binding site specific for the nucleic acid binding domain and 2) the recombinant binding moiety-encoding nucleic acids, wherein the recombinant binding moiety is expressed and binds to the binding site to form a vector- binding moiety complex; (8) lysing the host cells under conditions that do not disrupt binding of the vector-binding moiety complex; (9) contacting the vector-binding moiety complex with a target cell of interest and identifying target cells that contain the vector; and (10) isolating the optimized recombinant binding moiety nucleic acids from the target cells which contain the vector; and

(11) repeating (6) through (10), as necessary, to obtain a further optimized cell-specific binding moiety useful for increasing uptake or specificity of a genetic vaccine vector to a target cell.

The invention also provides cell-specific recombinant binding moieties produced by expressing in a host cell an optimized recombinant binding moiety-encoding nucleic acid obtained by the methods of the invention.

In another embodiment, the invention provides genetic vaccines that include: a) an optimized recombinant binding moiety that comprises a nucleic acid binding domain and a cell-specific ligand, and b) a polynucleotide sequence that comprises a binding site, wherein the nucleic acid binding domain is capable of specifically binding to the binding site.

A further embodiment of the invention provides methods for obtaining an optimized cell-specific binding moiety useful for increasing uptake, efficacy, or specificity of a genetic vaccine for a target cell by: reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid that comprises a polynucleotide which encodes a non-toxic receptor binding moiety-of an enterotoxin or other toxin, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids; (2) transfecting vectors that contain the library of nucleic acids into a population of host cells, wherein the nucleic acids are expressed to form recombinant cell- specific binding moiety polypeptides; (3) contacting the recombinant cell-specific binding moiety polypeptides with a cell surface receptor of a target cell; and (4) determining which recombinant cell-specific binding moiety polypeptides exhibit enhanced ability to bind to the target cell. Methods of enhancing uptake of a genetic vaccine vector by a target cell by coating the genetic vaccine vector with an optimized recombinant cell-specific binding moiety produced by these methods are also provided by the invention.

The present invention also provides methods for evolving a vaccine delivery vehicle, genetic vaccine vector, or a vector component to obtain an optimized delivery vehicle or component that has, or confers upon a vector, enhanced ability to enter a selected mammalian tissue upon administration to a mammal. These methods involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) members of a pool of polynucleotides to produce a library of experimentally generated (in vitro &/or in vivo) polynucleotides; (2) administering to a test animal a library of replicable genetic packages,

each of which comprises a member of the library of experimentally generated (in vitro &/or in vivo) polynucleotides operably linked to a polynucleotide that encodes a display polypeptide, wherein the experimentally generated (in vitro &/or in vivo) polynucleotide and the display polypeptide are expressed as a fusion protein which is which is displayed on the surface of the replicable genetic package; and (3) recovering replicable genetic packages that are present in the selected tissue of the test animal at a suitable time after administration, wherein recovered replicable genetic packages have enhanced ability to enter the selected mammalian tissue upon administration to the mammal.

If further optimization of the delivery vehicle is desired, the methods of the invention further involve: (4) reassembling (&/or subjecting to one or more directed evolution methods described herein) a nucleic acid that comprises at least one experimentally generated (in vitro &/or in vivo) polynucleotide obtained from a replicable genetic package recovered from the selected tissue with a further pool of polynucleotides to produce a further library of experimentally generated (in vitro &/or in vivo) polynucleotides; (5) administering to a test animal a library of replicable genetic packages, each of which comprises a member of the further library of experimentally generated (in vitro &/or in vivo) polynucleotides operably linked to a polynucleotide that encodes a display polypeptide, wherein the experimentally generated (in vitro &/or in vivo) polynucleotide and the display polypeptide are expressed as a fusion protein which is which is displayed on the surface of the replicable genetic package; (6) recovering replicable genetic packages that are present in the selected tissue of the test animal at a suitable time after administration; and (7) repeating (4) through (6), as necessary, to obtain a further optimized recombinant delivery vehicle that exhibits further enhanced ability to enter a selected mammalian tissue upon administration to a mammal. Methods of administration that are of particular interest include, for example, oral, topical, and inhalation. Where the administration is intravenous, mammalian tissues of interest include, for example, lymph node and spleen.

In another embodiment, the invention provides methods for evolving a vaccine delivery vehicle, genetic vaccine vector, or a vector component to obtain an optimized delivery vehicle or component to obtain an optimized delivery vehicle or vector component that has, or confers upon a vector containing the component, enhanced specificity for antigen-presenting cells by: reassembling (&/or subjecting to one or more directed evolution methods described herein) members of a pool of polynucleotides to produce a

library of experimentally generated (in vitro &/or in vivo) polynucleotides; producing a library of replicable genetic packages, each of which comprises a member of the library of experimentally generated (in vitro &/or in vivo) polynucleotides operably linked to a polynucleotide that encodes a display polypeptide, wherein the experimentally generated (in vitro &/or in vivo) polynucleotide and the display polypeptide are expressed as a fusion protein which is which is displayed on the surface of the replicable genetic package; (3) contacting the library of recombinant replicable genetic packages with a non-APC to remove replicable genetic packages that display non-APC-specific fusion polypeptides; and (4) contacting the recombinant replicable genetic packages that did not bind to the non-APC with an APC and recovering those that bind to the APC, wherein the recovered replicable genetic packages are capable of specifically binding to APCs.

In an additional embodiment, the invention provides methods for evolving a vaccine delivery vehicle, genetic vaccine vector, or a vector component to obtain an optimized delivery vehicle or component to obtain an optimized delivery vehicle or vector component that has, or confers upon a vector containing the component, an enhanced ability to enter a target cell by: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which encodes an invasin polypeptide, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant invasin nucleic acids; (2) producing a library of recombinant bacteriophage, each of which displays on the bacteriophage surface a fusion polypeptide encoded by a chimeric gene that comprises a recombinant invasin nucleic acid operably linked to a polynucleotide that encodes a display polypeptide; (3) contacting the library of recombinant bacteriophage with a population of target cells; (4) removing unbound phage and phage which is bound to the surface of the target cells; and (5) recovering phage which are present within the target cells, wherein the recovered phage are enriched for phage that have enhanced ability to enter the target cells.

In some embodiments, the optimized recombinant genetic vaccine vectors, delivery vehicles, or vector components obtained using these methods exhibit improved ability to enter an antigen presenting cell. These methods can involve washing the cells after the transfection step to remove vectors which did not enter an antigen presenting cell.; culturing the cells for a predetermined time after transfection; lysing the antigen presenting cells; and isolating the optimized recombinant genetic vaccine vector from the cell lysate.